

CLAIMS

1. A speech synthesis device, characterized in that the device comprises:

5 voice unit storage means that stores a plurality of voice unit data representing a voice unit;

selection means that inputs sentence information representing a sentence and selects voice unit data whose reading is common with a speech sound comprising
10 the sentence from the respective voice unit data;

missing part synthesis means that, for a speech sound among the speech sounds comprising the sentence for which the selection means could not select voice unit data, synthesizes speech data representing a
15 waveform of the speech sound; and

synthesis means that generates data representing synthetic speech by combining voice unit data that was selected by the selection means and speech data that was synthesized by the missing part synthesis means.

20

2. A speech synthesis device, characterized in that the device comprises:

voice unit storage means that stores a plurality of voice unit data representing a voice unit;

25 cadence prediction means that inputs sentence information representing a sentence and predicts the cadence of a speech sound comprising the sentence;

selection means that selects, from the respective voice unit data, voice unit data whose reading is common
30 with a speech sound comprising the sentence and whose

cadence matches a cadence prediction result under predetermined conditions;

missing part synthesis means that, for a speech sound among the speech sounds comprising the sentence
5 for which the selection means could not select voice unit data, synthesizes speech data representing a waveform of the voice unit; and

synthesis means that generates data representing synthetic speech by combining voice unit data that was
10 selected by the selection means and speech data that was synthesized by the missing part synthesis means.

3. The speech synthesis device according to claim 2, characterized in that the selection means excludes
15 from the objects of selection voice unit data whose cadence does not match a cadence prediction result under the predetermined conditions.

4. The speech synthesis device according to claim 20 2 or 3, characterized in that the missing part synthesis means comprises:

storage means that stores a plurality of data representing a phoneme or a phoneme fragment that comprises a phoneme; and

25 synthesis means that, by identifying phonemes included in the speech sound for which the selection means could not select voice unit data and acquiring from the storage means data representing the identified phonemes or phoneme fragments that comprise the phonemes
30 and combining these together, synthesizes speech data

representing a waveform of the speech sound.

5. The speech synthesis device according to claim 4, characterized in that the missing part synthesis means comprises missing part cadence prediction means that predicts the cadence of the speech sound for which the selection means could not select voice unit data, and

the synthesis means identifies a phoneme included in the speech sound for which the selection means could not select voice unit data and acquires from the storage means data representing the identified phoneme or a phoneme fragment that comprises the phoneme, converts the acquired data such that the phoneme or the phoneme fragment represented by the data matches the cadence result predicted by the missing part cadence prediction means, and combines the converted data together to synthesize speech data that represents the waveform of the speech sound.

20

6. The speech synthesis device according to claim 2, 3 or 4, characterized in that, for a speech sound for which the selection means could not select voice unit data, the missing part synthesis means synthesizes speech data representing the waveform of the voice unit based on the cadence predicted by the cadence prediction means.

7. The speech synthesis device according to any one of claims 2 to 6, characterized in that the voice

30

unit storage means associates cadence data representing time variations in a pitch of a voice unit represented by voice unit data with the voice unit data in question and stores the resulting data, and

5 the selection means selects, from the respective voice unit data, voice unit data whose reading is common with a speech sound comprising the sentence and for which a time variation in the pitch represented by the associated cadence data is closest to the cadence
10 prediction result.

8. The speech synthesis device according to any one of claims 1 to 7, characterized in that the device further comprises utterance speed conversion means that
15 acquires utterance speed data specifying conditions of a speed for producing the synthetic speech and selects or converts speech data and/or voice unit data comprising data representing the synthetic speech such that the speech data and/or voice unit data represents speech
20 that is produced at a speed fulfilling the conditions specified by the utterance speed data.

9. The speech synthesis device according to claim 8, characterized in that the utterance speed conversion
25 means, by eliminating a segment representing a phoneme fragment from speech data and/or voice unit data comprising data representing the synthetic speech or adding a segment representing a phoneme fragment to the voice unit data and/or speech data, converts the voice
30 unit data and/or speech data such that the voice unit

data and/or speech data represents speech that is produced at a speed fulfilling the conditions specified by the utterance speed data.

- 5 10. The speech synthesis device according to any one of claims 1 to 9, characterized in that the voice unit storage means associates phonetic data representing a reading of voice unit data with the voice unit data in question and stores the data, and
- 10 the selection means handles voice unit data which is associated with phonetic data representing a reading matching the reading of speech comprising the sentence as voice unit data whose reading is common with the speech.
- 15 11. A speech synthesis method, characterized in that the method comprises the steps of:
- storing a plurality of voice unit data representing a voice unit;
- 20 inputting sentence information representing a sentence;
- selecting voice unit data whose reading is common with a speech sound comprising the sentence from the respective voice unit data;
- 25 synthesizing speech data representing the waveform of a speech sound among the speech sounds comprising the sentence for which voice unit data could not be selected; and
- generating data representing synthetic speech by
- 30 combining voice unit data that was selected and speech

data that was synthesized.

12. A speech synthesis method, characterized in that the method comprises the steps of:

5 storing a plurality of voice unit data representing a voice unit;

inputting sentence information representing a sentence and predicting the cadence of speech sounds comprising the sentence;

10 selecting from the respective voice unit data, voice unit data whose reading is common with a speech sound comprising the sentence and whose cadence matches a cadence prediction result under predetermined conditions;

15 synthesizing speech data representing a waveform of a speech sound among the speech sounds comprising the sentence for which voice unit data could not be selected; and

generating data representing synthetic speech by
20 combining voice unit data that was selected and speech data that was synthesized.

13. A program for causing a computer to function as:

25 voice unit storage means that stores a plurality of voice unit data representing a voice unit;

selection means that inputs sentence information representing a sentence and selects voice unit data whose reading is common with a speech sound comprising
30 the sentence from the respective voice unit data;

missing part synthesis means that, for a speech sound among the speech sounds comprising the sentence for which the selection means could not select voice unit data, synthesizes speech data representing a waveform of the speech sound; and

synthesis means that generates data representing synthetic speech by combining the voice unit data that was selected by the selection means and the speech data that was synthesized by the missing part synthesis means.

10

14. A program for causing a computer to function as:

voice unit storage means that stores a plurality of voice unit data representing a voice unit;

15 cadence prediction means that inputs sentence information representing a sentence and predicts the cadence of a speech sound comprising the sentence;

selection means that selects, from the respective voice unit data, voice unit data whose reading is common with a speech sound comprising the sentence and whose cadence matches a cadence prediction result under predetermined conditions;

20 missing part synthesis means that, for a speech sound among the speech sounds comprising the sentence for which the selection means could not select voice unit data, synthesizes speech data representing a waveform of the speech sound; and

25 synthesis means that generates data representing synthetic speech by combining the voice unit data that was selected by the selection means and the speech data

30

that was synthesized by the missing part synthesis means.

15. A speech synthesis device, characterized in that the device comprises:

5 voice unit storage means that stores a plurality of voice unit data representing a voice unit;

cadence prediction means that inputs sentence information representing a sentence and predicts the cadence of a speech sound comprising the sentence;

10 selection means that selects, from the respective voice unit data, voice unit data whose reading is common with a speech sound comprising the sentence and whose cadence is closest to a cadence prediction result; and

synthesis means that generates data representing
15 synthetic speech by combining together the voice unit data that were selected.

16. The speech synthesis device according to claim 15, characterized in that the selection means
20 excludes from the objects of selection voice unit data whose cadence does not match the cadence prediction result under predetermined conditions.

17. The speech synthesis device according to
25 claim 15 or 16, characterized in that the device further comprises utterance speed conversion means that acquires utterance speed data that specifies conditions of a speed for producing the synthetic speech and selects or converts speech data and/or voice unit data comprising
30 data representing the synthetic speech such that the

speech data and/or voice unit data represents speech that is produced at a speed fulfilling the conditions specified by the utterance speed data.

5 18. The speech synthesis device according to claim 17, characterized in that the utterance speed conversion means, by eliminating segments representing phoneme fragments from speech data and/or voice unit data comprising data representing the synthetic speech
10 or adding segments representing phoneme fragments to the voice unit data and/or speech data, converts the voice unit data and/or speech data such that the voice unit data and/or speech data represents speech that is produced at a speed fulfilling the conditions specified
15 by the utterance speed data.

 19. The speech synthesis device according to any one of claims 15 to 18, characterized in that the voice unit storage means associates cadence data representing
20 time variations in a pitch of a voice unit represented by voice unit data with the voice unit data in question and stores the data; and

 the selection means selects from the respective voice unit data, voice unit data whose reading is common
25 with a speech sound comprising the sentence and for which time variations in a pitch represented by the associated cadence data are closest to the cadence prediction result.

30 20. The speech synthesis device according to any

one of claims 15 to 19, characterized in that the voice unit storage means associates phonetic data representing the reading of voice unit data with the voice unit data in question and stores the data, and

5 the selection means handles voice unit data which is associated with phonetic data representing a reading that matches the reading of a speech sound comprising the sentence as voice unit data whose reading is common with the speech sound.

10

21. A speech synthesis method, characterized in that the method comprises the steps of:

storing a plurality of voice unit data representing a voice unit;

15 inputting sentence information representing a sentence and predicting the cadence of speech sounds comprising the sentence;

selecting from the respective voice unit data, voice unit data whose reading is common with a speech
20 sound comprising the sentence and whose cadence is closest to the cadence prediction result; and

generating data representing synthetic speech by combining together the voice unit data that were selected.

25

22. A program for causing a computer to function as:

voice unit storage means that stores a plurality of voice unit data representing a voice unit;

30 cadence prediction means that inputs sentence

information representing a sentence and predicts the cadence of speech sounds comprising the sentence;

selection means that selects from the respective voice unit data, voice unit data whose reading is common
5 with a speech sound comprising the sentence and whose cadence is closest to the cadence prediction result; and

synthesis means that generates data representing synthetic speech by combining together the voice unit data that were selected.